Frequently Asked Questions about Teachscape *Focus*™ Calibration

1. What is Calibration and what is its purpose?

Calibration is a formative assessment exercise designed to confirm that observers are accurately applying Charlotte Danielson's Framework for Teaching (FfT) rubric. Observers are considered "calibrated" when they have demonstrated the ability to score classroom teaching practice using the FfT rubric with sufficient accuracy. The Calibration exercise requires observers to score two 15-minute, grade-band specific video segments on each of the eight components of the FfT rubric. Feedback provided includes: accuracy of scores compared to those of expert raters, the rationales for the scores assigned, and recommendations for improving accuracy if needed.

2. How is Calibration different from the Proficiency Assessment in Teachscape *Focus*?

The Proficiency Assessment measures all sub-skills that are part of the Teachscape *Focus* Observer Training, including 1) distinguishing between appropriate evidence and interpretation; 2) distinguishing between appropriate evidence and statements that are biased or suggest professional preferences; 3) assigning evidence to the appropriate FfT component; 4) recognizing evidence that has been misaligned with an FfT component; and 5) assigning an accurate score for each of the eight FfT components based on a set of evidence. Calibration is a shorter "check" on one aspect of proficiency—assigning an accurate score for each of the eight FfT components based on a set of evidence.

3. Why can't Calibration be used as the Proficiency Assessment for observers?

Calibration is not designed to be a summative assessment of observation skills. It cannot provide as comprehensive a measure of proficiency as the Proficiency Assessment in all of the sub-skills that contribute to observer accuracy in applying the FfT. Rather, Calibration is intended to be a formative exercise to provide observers with feedback regarding their current level of accuracy and steps that may be useful in maintaining or improving that accuracy. It provides districts with additional information regarding which observers may need additional support or remediation if the district has requested that data.

4. What specific skills does Calibration measure?

Calibration measures an observer's ability to apply the FfT to classroom teaching practice and assign accurate scores for each of the eight FfT components based on a set of evidence.

5. How long does Calibration take?

Calibration should take approximately one hour to complete. There are two 15-minute videos. Approximately 30 minutes might be allocated to watching and scoring each video. Additional steps recommended to improve accuracy, if needed, will require an additional commitment of time on an individual basis.

6. Who should engage in Calibration?

Calibration is designed for observers who have completed the Observer Training and Proficiency Assessment in the Framework for Teaching Proficiency System (FfTPS) or the Teachscape *Focus* Observation Training and Assessment System; however, the system administrator can determine which observers have access to Calibration.

7. Must I have passed the Proficiency Assessment in order to do Calibration exercises?

The requirements to pass the Proficiency Assessment and to do Calibration exercises are both local policy decisions. Each of these tools is provided as part of the Teachscape *Focus* system to support these policy requirements.

8. How often must I engage in Calibration?

Periodic Calibration is recommended, but the frequency of Calibration is a local policy decision. To support these policy requirements, multiple Calibration windows can be offered in the academic year. At least three windows during December/January, February/March, and April/May is recommended. These recommendations are based on feedback by district users regarding typical timelines for observer training and testing, as well as the usual timeframe for classroom observation cycles.

9. How can I prepare for Calibration?

Completing several Scoring Practice videos may serve as a good refresher if you feel one is needed. Revisiting Observer Training, particularly benchmarks and rangefinder videos, may also be a worthwhile exercise.

10. What happens if my performance on Calibration reveals that I am not scoring accurately?

Recommendations for improving accuracy will be provided as part of the feedback process. Ultimately, next steps will be a local policy decision. This also can be viewed as an opportunity to return to the training videos and review the components that were challenging to score.

11. How accurate must my Calibration scores be in order to be considered "calibrated"?

There are different categories of scoring accuracy that result from the Calibration exercise. Each of these categories receive feedback specific to the category.

The categories are:

"Demonstrating Scoring Accuracy": This is considered the "green" category. Of the 16 scores assigned during the Calibration exercise, at least 12 scores are in exact agreement with the expert scores, and there should be no more than 2 discrepant scores.

"Needs Remediation and Monitoring": This is considered the "red" category. Of the 16 scores assigned during the Calibration exercise, 8 or more are discrepant scores, and 4 or fewer scores are exact matches to the master scores.

"Needs Practice and Support": This is considered the "yellow" category. The scores assigned are any combination of exact, adjacent, and discrepant scores that fall between the green and the red categories.

12. What is the research base for the recommended cut points for the three Calibration results?

The Calibration exercise does not report "pass" or "no pass" results for which cut points, in the traditional sense, would be required. The type of formative feedback generated in this exercise depends on the performance level with which an observer's scoring accuracy is best aligned. The three performance levels—green, yellow, and red—were developed based on experts' opinions of what is acceptable scoring accuracy in practice. For example, the highest level (green) reflects the standard of scoring accuracy (75% exact or 89% exact + adjacent) observers might be expected to maintain given the nature of the scoring task, the length of the score scale, and the intended purpose of observation data.

13. What recommendations are provided for districts whose observers are in the yellow or red categories of Calibration?

Districts receive recommendations on the district-level Calibration reports for observers who fall into the yellow or red categories. The recommendations are advisory in nature, with the assumption that districts will determine what is feasible and appropriate given contextual factors such as evaluation timeline and resources. The recommendations range from "additional practice and support," which may mean revisiting training in specific modules and doing additional scoring practice, as well as engaging in discussion with other "calibrated" observers or coaches, to "needs remediation and monitoring," which may mean more substantial intervention (in addition to retraining and scoring practice), such as partnering with another observer for observations and then discussing the observation experience and assignment of scores. Since each district's structures and processes for support will vary, recommendations are at a general level and are designed to alert districts to individuals who may need monitoring during the observation process.

14. If an observer falls into the red category, can the observer immediately re-calibrate to try and move up to the green category?

If an observer falls into the red category, it indicates more serious issues that will not likely be resolved by taking another Calibration exercise. The red category indicates that the observer has drifted significantly from an accurate application of the FfT rubric, and that the observer should revisit Observer Training and Scoring Practice. It is also highly recommended that these observers are provided with additional support to regain their accuracy in assigning scores, and that their observations are monitored and/or suspended while this process is under way.